

CLAIMS

What is claimed is:

1. A processor-based method for analyzing contracts, comprising:
determining at least one language pattern indicative of a contract attribute from text of a plurality of contracts;
determining whether the language pattern is present in a contract; and
in response to the presence of the language pattern in the contract, assigning text associated with the language pattern to the contract attribute.
2. The method of claim 1, wherein determining at least one language pattern indicative of the contract attribute comprises identifying, from the plurality of contracts, annotations that describe a structural context associated with the language pattern.
3. The method of claim 2, further comprising manually adding the annotations to the plurality of contracts.
4. The method of claim 2, wherein the annotations comprise extensible markup language tags.
5. The method of claim 1, wherein the contract attribute is specified in a component object model associated with the contract.
6. The method of claim 1, wherein determining at least one language pattern indicative of the contract attribute comprises generating a rule having a structural context component associated with the contract attribute and a regular expression associated with the language pattern.
7. The method of claim 6, wherein the regular expression is formed using a top-down induction method.

8. The method of claim 6, wherein the structural context component is specified in a document object model associated with the contract.

9. The method of claim 6, wherein determining whether the language pattern is present in the contract further comprises classifying a portion of the contract containing the language pattern into a subject category associated with the structural context component of the rule.

10. The method of claim 9, wherein classifying the portion of the contract comprises classifying into the subject category based on at least one language pattern in the portion indicative of the subject category.

11. A system, comprising:

- a storage arrangement including a plurality of contracts stored in machine-readable form;

- a learning arrangement coupled to the storage arrangement and configured to determine at least one language pattern indicative of a contract attribute from text of the plurality of contracts;

- an extractor configured to determine whether the language pattern is present in a contract, the extractor further configured to, in response to the presence of the language pattern in the contract, assign a contract attribute to a portion of the text of the contract associated with the language pattern; and

- a contracts facts database configured to store a data value conforming to the portion of the text assigned to the contract attribute.

12. The system of claim 11, wherein the learning arrangement is configured to determine at least one language pattern indicative of the contract attribute by identifying, from the plurality of contracts, annotations that describe a structural context associated with the language pattern.

13. The system of claim 12, wherein the learning arrangement is configured to accept a user input for manually adding annotations to the plurality of contracts.

200310995-1 (HPCO.139PA)

14. The system of claim 12, wherein the annotations comprise extensible markup language tags.

15. The system of claim 11, wherein the learning arrangement is configured to determine at least one language pattern indicative of the contract attribute by generating a rule having a structural context component associated with the contract attribute and a regular expression associated with the language pattern.

16. The system of claim 15, wherein the rule is generated using a top-down induction method to form the regular expression.

17. The system of claim 11, wherein the contracts database comprises a relational database.

18. The system of claim 11, wherein the contracts database comprises an extensible markup language database.

19. A computer-readable medium configured with instructions for causing a processor of a data processing arrangement to perform steps comprising:
determining at least one language pattern indicative of a contract attribute from text from a plurality of contracts;
determining whether the language pattern is present in a contract; and
in response to the presence of the language pattern in the contract, assigning a portion of text associated with the language pattern to the contract attribute.

20. The computer-readable medium of claim 19, wherein determining at least one language pattern indicative of the contract attribute comprises identifying, from the plurality of contracts, annotations that describe a structural context associated with the language pattern.

21. The computer-readable medium of claim 20, wherein the steps further comprise manually adding the annotations to the plurality of contracts.

200310995-1 (HPCO.139PA)

22. The computer-readable medium of claim 20, wherein the annotations comprise extensible markup language tags.

23. The computer-readable medium of claim 19, wherein determining at least one language pattern indicative of the contract attribute comprises generating a rule having a structural context component associated with the contract attribute and a regular expression associated with the language pattern.

24. The computer-readable medium of claim 23, wherein the rule is generated using a top-down induction method to form the regular expression.

25. A system comprising:

means for determining at least one language pattern indicative of a contract attribute from text from a plurality of contracts;

means for determining whether the language pattern is present in a contract;
and

means for assigning text of the contract to a contract attribute in response to the presence of the language pattern in the contract.

26. The system of claim 25, further comprising means for identifying, from the plurality of contracts, annotations that describe a structural context associated with the language pattern.